

RESEARCH PAPER



DNA methylation differentiates smoking from vaping and non-combustible tobacco use

Allan Andersen^a, Rachel Reimer^b, Kelsey Dawes^a, Ashley Becker^a, Natasha Hutchens^a, Shelly Miller^c, Meesha Dogan^{a,c,d}, Brandon Hundley^a, James A Mills^a, Jeffrey D Long^{a,e}, and Robert Philibert^{a,c,d}

^aDepartment of Psychiatry, University of Iowa, Iowa City, USA; ^bCollege of Public Health, Des Moines University, Des Moines, USA; ^cBehavioral Diagnostics LLC, Coralville, USA; ^dRoy J. Carver Department of Biomedical Engineering, University of Iowa, Iowa City, United States;

^eDepartment of Biostatistics, University of Iowa, Iowa City, USA

ABSTRACT

Increasing use of non-combusted forms of nicotine such as e-cigarettes poses important public health questions regarding their specific risks relative to combusted tobacco products such as cigarettes. To fully delineate these risks, improved biomarkers that can distinguish between these forms of nicotine use are needed. Prior work has suggested that methylation status at cg05575921 may serve as a specific biomarker of combusted tobacco smoke exposure. We hypothesized combining this epigenetic biomarker with conventional metabolite assays could classify the type of nicotine product consumption. Therefore, we determined DNA methylation and serum cotinine values in samples from 112 smokers, 35 e-cigarette users, 19 smokeless tobacco users, and 269 controls, and performed mass spectroscopy analyses of urine samples from all nicotine users and 22 verified controls to determine urinary levels of putatively nicotine product-specific substances; propylene glycol, 2-cyanoethylmercapturic acid (CEMA), and anabasine. 1) Cigarette smoking was associated with a dose dependent demethylation of cg05575921 and increased urinary CEMA and anabasine levels, 2) e-cigarette use did not demethylate cg05575921, 3) smokeless tobacco use also did not demethylate cg05575921 but was positively associated with anabasine levels 4) CEMA and cg05575921 levels were highly correlated and 5) propylene glycol levels did not reliably distinguish use groups. Cg05575921 assessments distinguish exposure to tobacco smoke from smokeless sources of nicotine including e-cigarettes and smokeless tobacco, neither of which are associated with cg05575921 demethylation. A combination of methylomic and metabolite profiling may allow for accurate classification use status of a variety of nicotine containing products.

ARTICLE HISTORY

Received 21 October 2020

Revised 8 January 2021

Accepted 10 February 2021

KEYWORDS

AHRR; DNA methylation; cg05575921; anabasine; 2-cyanoethylmercapturic acid; propylene glycol; vaping; smoking


Introduction

The past several years have seen a dramatic rise in the consumption of nicotine using electronic nicotine delivery systems (ENDS) such as e-cigarettes and ‘vape pens.’ In particular, the use of ENDS by adolescents has surpassed the use of more traditional forms of nicotine-containing products such as cigarettes and chewing tobacco in this age group [1]. According to one large national study, 25% of high school seniors reported ‘vaping’ nicotine or flavour containing fluids in 2018[2].

In general, the concerns that the healthcare community has with respect to the use of ENDS are at least five-fold. First, there are concerns about the acute and long-term effects of nicotine itself[3]. Second, there are grave concerns that the

use of these non-combustible nicotine delivery vehicles may increase the likelihood of smoking or other forms of substance use[4]. Third, the long-term effects of inhalation of atomized glycerol or propylene glycol, the two main solvents used in ENDS, may have deleterious effects on lung function[5]. Fourth, as evidenced by a recent plague of acute pulmonary injuries attributed to the inclusion of vitamin E acetate in some vaping solutions, adulterants in the vaping fluid can have catastrophic healthcare consequences[6]. Fifth and finally, toxic heavy metals, such as cadmium, can be released from the heating coils of these devices[7]. Given the fact that use of ENDS has been strongly embraced by many adolescents and young adults, it is unlikely that these devices

CONTACT Allan Andersen  allan-andersen@uiowa.edu Department of Psychiatry, University of Iowa, Iowa City, IA 52242, USA

 Supplemental data for this article can be accessed [here](#).

© 2021 Informa UK Limited, trading as Taylor & Francis Group

will disappear from the market. Therefore, it is incumbent upon the healthcare community to develop a comprehensive understanding of each of these potential risks so that policymakers can best focus the available societal resources to moderate their use.

Although constraining the extent of each of these five effects of the use of ENDS is important, defining the relationship of ENDS use to smoking and to other substance use may be the most critical. Smoking leads to the premature death of nearly ½ million Americans each year[8]. Prior to the onset of the vaping epidemic, the rate of adult smoking was at a historic low[9]. If ENDS use facilitates the initiation of smoking, this new form of nicotine consumption may undo the work of countless healthcare and public health officials by unleashing a new wave of adult smokers.

A major challenge to developing a better understanding of the relationship between ENDS use is the inability of the most traditionally trusted biochemical means for ascertaining smoking status, the cotinine test, to differentiate between smoking and ENDS use. Nicotine, regardless of whether it comes from vaping, smoking cigarettes, non-combustible tobacco, or even the nicotine patch, produces the same indistinguishable form of cotinine[10]. As a consequence, each of these types of nicotine use results in a positive cotinine test. Conceivably, by using more complex mass spectroscopy methods, it is possible to detect other tobacco alkaloids and provide a means to differentiate smokers from vapers [11]. But these methods are expensive, time-consuming, and only assess consumption over the past 24 to 48 hours[12].

One promising alternative method for differentiating smokers from vapers may be through the use of DNA methylation. Over the past 5 years, dozens of studies have shown that smoking is associated with genome-wide changes in DNA methylation [13,14]. In particular, these studies have shown that DNA methylation status at cg05575921, a CpG residue in the Aryl Hydrocarbon Receptor Repressor (AHRR), is highly predictive of smoking status. Demethylation of cg05575921 results in an increased expression of AHRR, which is a key feedback regulator in the Xenobiotic pathway [15–17].

Critically, because the AHRR gene is a feedback regulator of CYP1A1, CYP1A2 and CYP1B1, each

of which are cytochromes that moderate the metabolism of the polycyclic aromatic hydrocarbons in smoke [17], but not the cytochromes moderating nicotine metabolism (mainly CYP2A6) [18], methylation status at this locus should not be affected by ingestion of nicotine products that are not combusted, as in ENDS. Indeed, a 2013 genome-wide methylation study of 74 ‘snuff’ (a pulverized form of tobacco that can be placed in the mouth or snorted through the nose used commonly in Scandinavia [19]) users and 378 controls by Besingi and Johansson failed to show any significant changes in whole-blood DNA methylation, in particular at AHRR, as compared to controls associated with snuff use[20]. Similarly, we found no change in cg05575921 methylation associated with the two users of chewing tobacco[21]. If these findings that show no effect of non-combustible tobacco use generalize to other non-pyrolysis methods of ingesting nicotine, this suggests the possibility that cg05575921 methylation may be a mechanism through which to specifically determine smoking status for both clinical and research applications.

A barrier to the potential implementation of using DNA methylation to quantitate smoking is the costly, time-consuming nature of measuring DNA methylation using genome-wide arrays. In an effort to speed up the transition of basic epigenetic findings into usable clinical tools, we have translated the bulky, time-consuming array-based methods for assessing DNA methylation into a methylation sensitive digital PCR assay capable of being rapidly performed in any laboratory possessing digital PCR equipment[22]. Studies by our group have shown that this assay can accurately predict cigarette smoking in both adolescents and adults using DNA prepared from either blood or saliva [23–25].

Theoretically, if this methylation assay were used in tandem with a method for detecting nicotine ingestion, such as serum, salivary or urinary cotinine assay, it should be possible to quickly differentiate those who only vape nicotine containing fluids from those who vape and smoke by determining both cg05575921 and cotinine status. However, rigorous and well-designed studies to sensitively and specifically test this hypothesis have not yet been reported.

In this communication, we use the digital PCR assay in combination with standard serum cotinine measures and state of the art mass spectroscopy to determine whether a combination of methylomic and metabolomic methods can effectively predict cigarette, vaping and smokeless tobacco consumption in individuals carefully screened into each use group.

Methods

Ethics: All protocols and procedures used in this study were approved by the University of Iowa Institutional Review Board (IRB 201,905,678). For this study, the University of Iowa has agreed to serve as the lead IRB for both the Iowa and the Des Moines University activities.

Subjects: Potential participants were solicited through mass email to the University of Iowa and Des Moines University community, social media ads, radio ads, and fliers. Each form of recruitment referred interested individuals to a University of Iowa website hosting a RedCap pre-screening interview[26]. The pre-screening instrument, which was conducted anonymously, is an algorithm containing 50 interrogatives that attempted to determine whether the potential subject was over the age of 18, capable of giving informed consent in English, and fits into one of four nicotine product use categories. ‘Controls’ were defined by this algorithm as individuals who denied smoking more than 100 cigarettes or joints in their lifetime, denied the use of cannabis products in the past year, had not used any nicotine containing products in the past year, denied a history of substance use problems, and had consumed less than 2 alcoholic drinks per day. ‘Cigarette Smokers’ were defined as individuals who have at least 5 pack years of consumption, currently smoke two or more cigarettes per day, but do not currently use vaping products. ‘Vapers’ were defined as those who use e-cigarettes or ‘vape’ daily, have ‘vaped’ at least once per week for the past year, deny smoking more than 100 cigarettes in their lifetime, no use of cannabis for at least 1 year, and no other use of tobacco containing products. ‘Smokeless users’ were defined as individuals who currently use chewing tobacco on a daily basis, denied smoking more than 100

cigarettes in their lifetime, no use of cannabis for at least 1 year, no other use of tobacco containing products, and no vaping. Individuals fitting the inclusion and exclusion criteria for one of these four groups were then given an opportunity to schedule themselves for a more comprehensive in-person interview. Subjects who did not qualify for the main study were notified of their ineligibility at the completion of the screening interview.

At the time of the in-person intake interview, the potential subject was given a written summary of the project and given a chance to ask questions. If still willing to participate, full written consent was obtained, and the study procedures initiated. The clinical interview consisted of a 320-item interview administered through a RedCap interface comprised of questions from the PhenX Toolkit [27] that detailed demographic and substance use histories of each individual. Self-reported cigarette consumption was assessed the following question; ‘*How frequently have you smoked cigarettes during the past (day/week etc.)?*’ Similarly, e-cigarette use was assessed by asking ‘*How frequently have you used e-cigarettes during the past (day/week etc.)?*’ Finally, smokeless tobacco consumption was assessed by asking ‘*How frequently have you used a smokeless tobacco product during the past (day/week etc.)?*’ For each of these questions, the participants selected one of seven potential consumption levels (e.g., “two to five cigarettes per day”). A complete copy of the interview is available upon request.

Although the interview itself was computer administered, a research assistant remained present to answer any questions. After the interview was complete, the subject then provided urine and then was phlebotomized.

Laboratory measures

The blood was processed into DNA and serum via our usual procedures and then stored at -20°C and -80°C , respectively. Urine was apportioned into 20 ml aliquots and then stored at -20°C until analysis.

Exhaled carbon monoxide (CO) levels were assessed using a Smokelyzer[®] according to manufacturer’s directions (CoVita, USA).

Enzyme linked immunoassay (ELISA) assessments of serum cotinine and tetrahydrocannabinol (THC) levels were conducted using kits from AbNova (Taiwan) using our standard procedures [24,28]. Based on the results of these tests, the data from 14 self-reported non-smoking control subjects were excluded for having serum cotinine levels >2 ng/ml. Eight self-reported daily cigarette smokers were excluded because their serum cotinine levels were <2 ng/ml. Data from six other subjects who enrolled in the study were excluded for not providing smoking use history in their intake interview (n = 2) or for significant discrepancies in the nicotine product use histories reported in the screening instrument as compared to the intake interviews (n = 6).

Secondary to the discontinuation of production of the ELISA kits during the course of this study, only 59 of the serum samples from the control subjects were examined for THC positivity. However, sera from all subjects from the smoking, vaping and smokeless groups were tested for THC positivity.

The determination of DNA methylation at cg05575921 was conducted using universal droplet digital PCR reagents and equipment from Bio-Rad (Hercules, CA) and a proprietary primer probe set from Behavioral Diagnostics (Coralville, IA) according to our previously described protocols [13,24]. The resulting DNA methylation levels are reported as percent methylation (Methyl CpG/Total CpG).

In order to assess the accuracy of self-report of nicotine product use status, we determined spot urinary levels of cotinine, a well-characterized metabolite of nicotine; 2-cyanoethylmercapturic acid (CEMA), a urinary metabolite of acetonitrile whose presence is predictive of smoking[29]; propylene glycol (PG), which is a solvent used in vaping solution; and two tobacco alkaloids, anabasine and anatabine [30], in samples from all Smokers, Vapers, and Smokeless users, as well as a random sample of 22 controls. These analyses were conducted under contract by Analytisch-Biologisches Forschungslabor (ABF) Laboratories (Munich, <https://abf-lab.com/>) using proprietary high pressure liquid chromatography (HPLC) or gas chromatography (GC) mass spectroscopy procedures.

Data Analysis: Data were analysed using SAS 9.4 (SAS Institute, Cary, SC) and R version 4.0.2. Comparisons between groups with respect to continuous variables were conducted using ANCOVA with Tukey-Kramer Honest Significant Difference (HSD) used to compare significance between groups[31]. Linear and cubic spline regression was used to analyse the relationship between continuous variables with model comparisons made using AIC values [32,33]. All reported R² values are adjusted for the number of predictors and sample size. Logistic regression was used to predict smoking status (i.e., smoker vs. control) as indicated by serum cotinine level and to predict use type (i.e., smoker vs. vaper/smokeless) from a combination of exhaled carbon monoxide (CO), cg05575921, and CEMA. The strength of the effect was determined using the Receiver Operator Characteristic (ROC) area under the curve (AUC) derived from the logistic regression results [34,35].

Data Availability: All data used in this study are available from the corresponding author on reasonable request.

Results

(Table 1) lists the key demographic and clinical characteristics of the 435 subjects included in the main analyses. Consistent with the demographics of the state of Iowa, the subjects in each of the groups were generally white. However, the groups did differ in age and gender. The Vaping subjects (n = 35) were significantly younger than the smoking (n = 112, p < 0.0001), smokeless (n = 19, p < 0.001), and the control (n = 269, p < 0.01) groups. Conversely, the Smokeless tobacco subjects were almost exclusively male (18 of 19), whereas the other groups were preponderantly female.

There were significant differences between groups in other key demographic characteristics as well. Consistent with their comparatively younger ages, most of the vaping subjects (83%) reported never having married. In contrast, the older control (43%) smokeless (58%), and smoking (54%) groups tended to be in committed (married or co-habiting) relationships. Similarly, there were significant differences in family income and

Table 1. Clinical and demographic characteristics of the subjects.

	Control	Control subgroup	Smoker	Vaper	Smokeless
N	269	22	112	35	19
Age	30.4 ± 11.8	33.5 ± 12.2	41.2 ± 12.9	23.5 ± 4.9	36.6 ± 12.3
Race					
White	232	17	92	29	18
African American	6	1	10	2	0
Asian	14	2	3	1	0
Other	7	2	1	0	0
More than one	7	0	5	3	1
Ethnicity					
Hispanic	11	0	3	3	0
Non-Hispanic	257	22	109	32	19
Gender					
Male	80	3	45	13	18
Female	187	10	67	22	1
Other	1				
Marital status					
Single	139	9	35	29	6
Married or cohabiting	115	12	60	6	11
Divorced or separated	12	1	16	0	1
Education					
<12 years	0	0	2	0	0
HS diploma or GED	1	1	2	1	0
Some college	10	4	34	5	1
Bachelor's degree	71	9	32	17	3
Some post-baccalaureate	166	3	40	12	15
Doctoral or professional	21	5	2	0	0
Family income (in \$1000)	\$98 ± 95	\$91 ± 98	\$61 ± 75	\$79 ± 62	\$83 ± 48

education. The Control subjects had significantly greater income than the smoking group ($p < 0.05$). Control subjects were also more likely to be better educated than the smoking but not the smokeless subjects.

(Table 2) lists the key self-report and objective use data for nicotine products cannabis and alcohol for each of the subjects. Smokers reported consuming approximately 15 cigarettes per day throughout each of the time periods queried. Vaping subjects used their devices an average of 3 times per day through the same set of time periods. Finally, the smokeless

tobacco users reported an average of 3–4 instances of chew use each day.

Since nicotine is the primary driver of dependency for tobacco use disorders, we first compared the serum values of the most easily detected nicotine metabolite, cotinine. Overall, the ANCOVA analysis of the serum cotinine data showed that after adjustment for age and gender, the smokeless tobacco users had significantly higher cotinine values than the smokers and vapers ($p < 0.05$ and $p < 0.001$, HSD, respectively) while smokers and vapers did not differ ($p < 0.15$, HSD).

Table 2. Subjective and objective nicotine product use measures.

	Controls	Control subgroup	Smokers	Vapers	Smokeless
N	269	22	112	35	19
Unit of consumption			'Cigarettes'	'Use'	'Product'
Past day	-	-	13.6 ± 8.9	2.7 ± 1.0	3.2 ± 0.8
Past week	-	-	15.6 ± 8.8	2.9 ± 0.8	3.4 ± 0.7
Past month	-	-	15.2 ± 8.7	3.0 ± 0.8	3.3 ± 0.8
Past 6 months	-	-	15.5 ± 8.5	3.0 ± 0.8	3.4 ± 0.7
Past year	-	-	15.6 ± 8.6	3.0 ± 0.9	3.2 ± 0.9
Pack year history			13.1 ± 13.1		
Average WB cg05575921	86.8% ± 2.8	86.8% ± 2.1	54.6% ± 19.2	84.0% ± 5.0	84.3% ± 4.6
Average exhaled CO (ppm)	0.8 ± 1.2	0.6 ± 0.9	12.3 ± 10.5	0.2 ± 0.5	0.6 ± 0.8
Serum cotinine (ng/ml)	-	-	92 ± 34	78 ± 45	108 ± 39
Use of cannabis in past year?	17	3	53	13	3
Positive serum THC?	0	0	37	6	1
Alcohol consumption					
Past year (drinks per week)	3.5 ± 2.5	3.6 ± 3.2	5.3 ± 4.7	6.2 ± 5.1	8.1 ± 5.2
Use occasions in past year	19 ± 15	22 ± 16	19 ± 16	22 ± 15	29 ± 15
Binge drinking past 2 weeks	0.4 ± 1.0	0.1 ± 0.3	0.7 ± 1.7	1.4 ± 1.8	1.7 ± 3.0

We then examined the relationship of the three biomarkers for cigarette use, cotinine, exhaled carbon monoxide (CO) and cg05575921 methylation to one another using just the data from the control and smoking groups. Figure 1 illustrates the relationship of serum cotinine levels to cg05575921 and CO levels. Both were highly correlated with serum cotinine values with linear models for both cg05575921 ($n = 366$, $\text{adj-R}^2 = 0.68$, $p < 0.0001$) and CO ($n = 368$, $\text{adj-R}^2 = 0.60$, $p < 0.0001$) producing excellent fits. Similarly, CO was highly correlated with cg05575921 ($n = 354$, $\text{adj-R}^2 = 0.54$, $p < 0.0001$; plot not shown).

We then modelled the relationship of daily cigarette consumption over the past week with each of the biomarkers. Using simple linear curves to model the dose response relationship produced excellent fits, with self-reported cigarette consumption being highly correlated with serum cotinine ($n = 381$, $\text{adj-R}^2 = 0.70$), cg05575921 ($n = 366$, $\text{adj-R}^2 = 0.62$) and exhaled CO ($n = 368$, $\text{adj-R}^2 = 0.50$) levels. Since both cotinine levels and cg05575921 methylation assess saturable metabolic pathways with non-linear kinetics [36,37], we next modelled their relationship to cigarette consumption using a cubic spline. Figure 2 shows the results of those analyses. The cubic spline fit markedly improved prediction of biomarker levels for both serum cotinine ($n = 381$, $\text{adj-R}^2 = 0.85$; $\Delta\text{AIC} = 267.2$) and cg05575921 ($n = 366$, $\text{adj-R}^2 = 0.71$; $\Delta\text{AIC} = 91.7$) levels. In

contrast, the cubic spline fit of exhaled CO levels produced a more modest improvement in fit ($n = 368$, $\text{adj-R}^2 = 0.54$; $\Delta\text{AIC} = 33.0$) as compared to the simple linear model.

Because of its relative stability, DNA methylation has also been proposed as a marker of lifetime tobacco consumption. Nevertheless, our analysis of the relationship of lifetime (pack year) cigarette consumption to cg05575921 methylation demonstrated only a modest correspondence ($n = 361$, $\text{adj-R}^2 = 0.10$, $p < 0.001$). Even weaker relationships were seen for the relationship of lifetime consumption with serum cotinine ($n = 376$, $\text{adj-R}^2 = 0.05$, $p < 0.02$) and exhaled carbon monoxide ($n = 363$, $\text{adj-R}^2 = 0.06$, $p < 0.01$) levels.

To explore the potential effect of different levels of vaping intensity on smoking biomarker levels, similar linear regression models including vaping frequency over the past 24 hours, past week, past month, past 6 months, and past year as predictors of serum cotinine, CO, and cg05575921 methylation were constructed. Analysis of these models indicated that vaping frequency over each time period was unrelated (all $p > 0.05$) to cg05575921 methylation and CO and was only related to serum cotinine within the 24 hours prior to study participation ($n = 35$, $\text{adj-R}^2 = 0.09$, $p < 0.05$).

In the final set of analyses that focused only on the control and smoking subjects, we compared the ability of each of the three biomarkers for

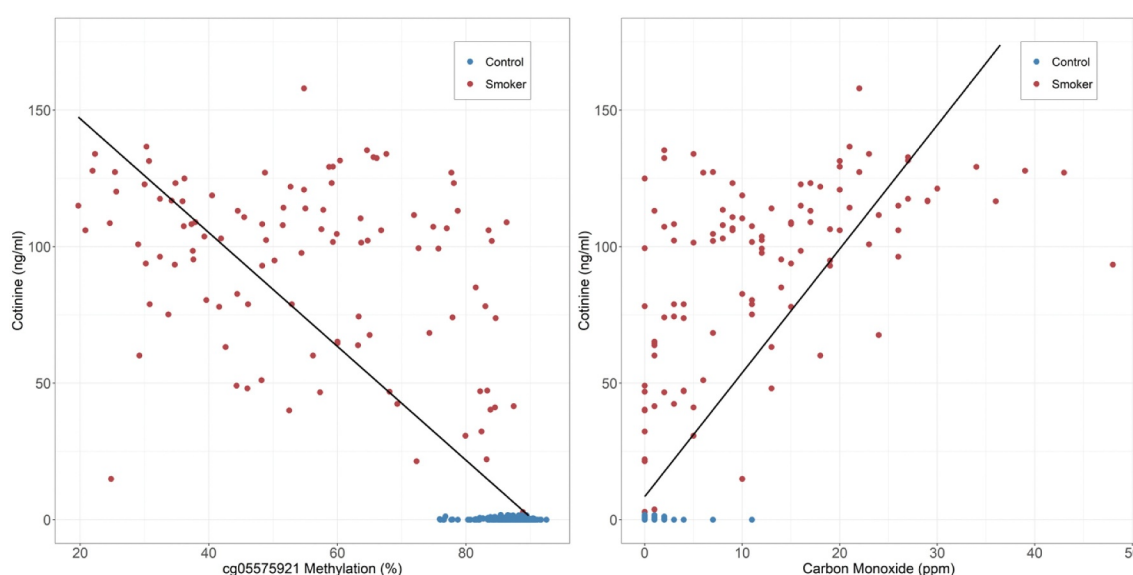


Figure 1. The relationship of serum cotinine and carbon monoxide levels for the control and smoking participants ($n = 366$, $\text{adj-R}^2 = 0.68$ for cg05575921 and $n = 368$, $\text{adj-R}^2 = 0.60$ for exhaled carbon monoxide).

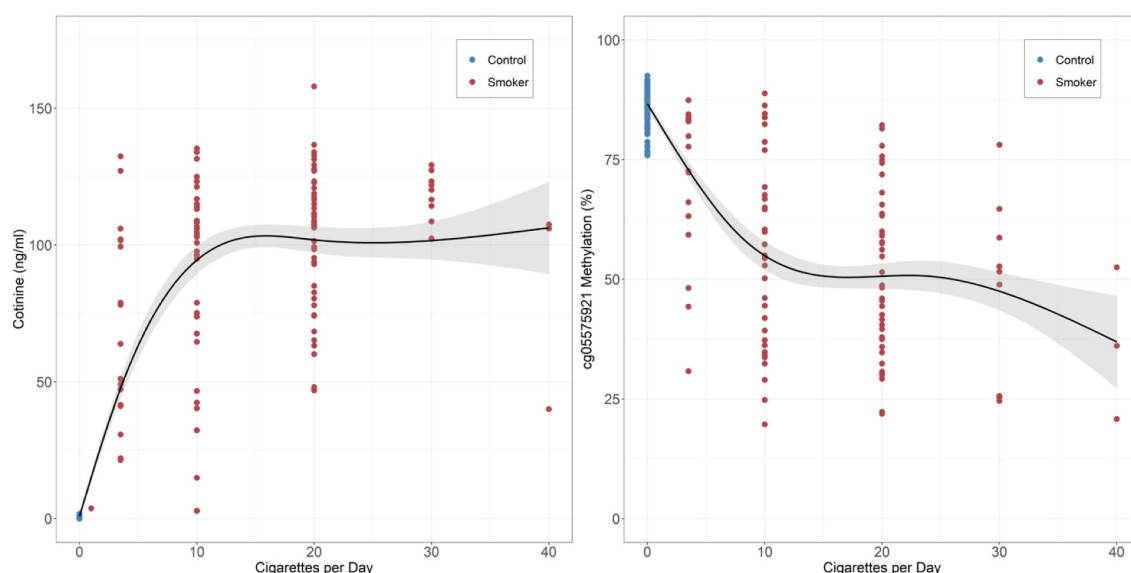


Figure 2. Cubic spline fit of the relationship of self-reported daily cigarette consumption over the past week to serum cotinine and cg05575921 levels for the control and smoking participants ($n = 381$, $\text{adj-R}^2 = 0.85$ and $n = 366$, $\text{adj-R}^2 = 0.71$, respectively).

smoking, serum cotinine, exhaled carbon monoxide and cg05575921 to predict smoking status using a standard receiver-operating characteristic (ROC) area under the curve (AUC) approach. Not surprisingly, since we excluded any control subject who had a serum cotinine of 2 ng/ μ l or a smoking subject with a serum cotinine of less than 2 ng/ μ l, serum cotinine levels perfectly predicted smoking status ($n = 381$). Cg05575921 predicted smoking status well with an AUC of 0.97 that was not improved by the addition of age or sex ($n = 366$). Similarly, exhaled CO also performed well with an AUC of 0.89 ($n = 368$). The addition of age to CO, which was significantly associated ($p < 0.003$) with smoking status, improved prediction slightly (AUC = 0.90). Nevertheless, even with the addition of age, a comparison of ROC curves shows that exhaled CO did not distinguish the two groups as well as cg05575921 methylation ($p < 0.005$).

We then expanded our analyses to include data from all four nicotine use groups. In line with our

previous effort to identify control or smoking subjects who may be providing unreliable data, we determined putative biomarkers specific for each tobacco use group, propylene glycol, cotinine, CEMA, anabasine and anatabine levels in urine specimens from almost every subject in each of the three user groups (smokers $n = 112$, vapers $n = 35$ and smokeless $n = 18$) and in a random sampling of control subjects ('control subgroup', $n = 22$) using state-of-the-art mass spectroscopy techniques.

(Table 3) presents the group averages and standard deviations for each group for each of the five analytes while Supplemental (Figure 1) visually summarizes the findings from those analyses for the four tobacco-related metabolites that showed interesting variance. Whether or not the effects of two unusually high reading (743 μ g/ml for one smoker and 305 μ g/ml for one vaping subject) are considered, there were no significant differences between any of the groups with respect to urinary levels of propylene glycol (Table 3).

Table 3. Results of mass spectroscopy analyses of product metabolites in urine.

	Control subgroup	Smokers	Vapers	Smokeless
N =	22	112	35	18
Propylene glycol (ug/ml)	4.1 ± 7.8	17.1 ± 71.5	19.4 ± 50.8	6.6 ± 8.5
Cotinine (ng/ml)	0.0 ± 0.1	1041 ± 769	777 ± 904	1598 ± 1504
CEMA (ng/ml)	1.1 ± 1.0	187.0 ± 204.4	4.1 ± 11.6	2.0 ± 4.5
Anabasine (ng/ml)	0.5 ± 2.0	10.5 ± 14.7	3.6 ± 9.9	14.8 ± 22.1
Anatabine (ng/ml)	0.4 ± 1.5	12.1 ± 17.3	1.9 ± 5.2	21.3 ± 33.1

Urinary levels of cotinine in the control group were essentially zero, with the average level of cotinine seen in smokeless tobacco users (1597 ng/ml) being significantly higher than that for vapers (777 ng/ml, $p < 0.05$, HSD) but not smokers (1041 ng/ml; Supplemental Figure 1). Levels of CEMA, which is produced by pyrolysis of tobacco, were significantly higher in smokers than all other groups (Supplemental Figure 1). Closer inspection of the five vaping subjects with CEMA levels of greater than 3.0 ng/ml showed that three of these subjects had cg05575921 levels below the 80% level associated with smoking. Since the control average for CEMA was 1.1 ± 1.0 ng/ml, this suggests that at least some of these five subjects may have been surreptitiously smoking. Finally, consistent with having their origin in tobacco, both anabasine and anatabine levels were significantly elevated in smokers and smokeless use groups as compared to Controls (both $p < 0.05$, HSD; Supplemental Figure 1). However, while arithmetically higher in vapers, neither anabasine nor anatabine were significantly

elevated in vapers as compared to controls. Interestingly, once again, the vaping subjects with higher levels of anabasine tended to have lower levels of cg05575921 methylation.

We next compared the relationship of serum cotinine to their paired urine specimen cotinine values. Although the urine levels were an order of magnitude higher than the corresponding serum values, the two sets of cotinine values were highly correlated with a simple linear model being highly significant ($\text{adj-}R^2 = 0.46$, $n = 187$, see Figure 3).

In an attempt to determine how well cotinine values predicted consumption in the vaping and smokeless subjects, we analysed serum and urinary cotinine values for the prior self-reported consumption for both of these groups. In general, visual inspection of the plots for the smokeless tobacco users shows good correspondence between the past day consumption report and both serum cotinine and urine cotinine (Supplementary Figure 2). However, a similar plotting of the 35 Vaping subjects failed to reveal a clear relationship between the number of times

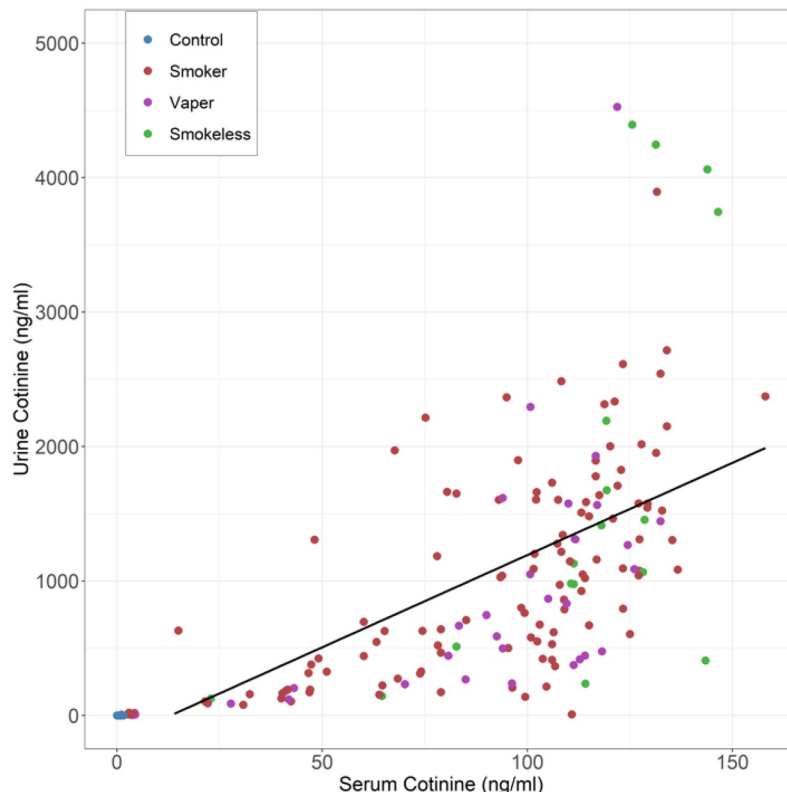


Figure 3. The relationship of serum cotinine to urine nicotine levels. The black line is the best fit line for a linear model ($N = 187$, $\text{adj-}R^2 = 0.46$).

that the subject reported vaping and the serum or urine cotinine values (Supplementary Figure 3). Still, in this regard, we will note that although to be included in this study, the vaping subjects had to report daily vaping but not daily vaping of nicotine containing products. Therefore, in these cases, the absence of serum or urine cotinine should not be construed to be indicative of unreliable self-report.

Finally, we ran a series of logistic regression models predicting smoking vs. vaping/smokeless use to determine if methylation status, in combination with CO and CEMA, could be used to classify the smoking group accurately. Starting with a base model including age and gender, we added cg05575921, CO, and CEMA individually, two at a time, and simultaneously. Table 4 shows Model 4 with CEMA was the best fitting model with a single predictor and Model 6 with CEMA and cg05575921 was the best fitting with two predictors. Model 8, with all three predictors, had the highest AUC and the lowest AIC (AUC = 0.990, AIC = 47.65, n = 146).

Discussion

In summary, we demonstrate that, in carefully pre-screened use groups: 1) cg05575921 methylation status can be used to reliably classify those who smoke from those who use non-combustible nicotine containing products and 2) a combination of methylation and metabolite testing can be used to classify users of the major forms of nicotine-related products. Finally, we confirm and extend our previous findings on the dose dependency of cg05575921 demethylation in response to smoking.

Table 4. AUC and AIC for logistic regression models predicting smoking vs. vaping/smokeless use.

Model [†]	Predictors	AUC	AIC
1	Age, gender	0.822	148.02
2	Age, gender, cg05575921	0.947	90.58
3	Age, gender, CO	0.950	81.37
4	Age, gender, CEMA	0.984	62.54
5	Age, gender, cg05575921, CO	0.974	63.78
6	Age, gender, cg05575921, CEMA	0.990	52.40
7	Age, gender, CO, CEMA	0.987	53.58
8	Age, gender, cg05575921, CO, CEMA	0.990	47.65

[†]Missing values for predictors result in analysis sample size of 146 participants for all models

The dose dependency of the demethylation response to smoking is very similar to that shown in our previous examination of daily smokers drawn from substance use treatment centres [24]. Still, the current sample, which is drawn from the general community and is significantly younger than the prior sample of smokers, has a substantially greater proportion of lighter smokers than our previous studies [21,24]. This additional information should help constrain the overall confidence intervals of cg05575921 in particular, at lower levels of cigarette consumption not well covered in previous analyses.

These quantitative data further emphasize the fact that smokers are not a homogenous group and suggest the possibility that the level of nicotine dependence in smokers may vary as a function of consumption levels. ‘Nondaily’ and ‘Social Smokers’ tend to smoke lower numbers of cigarettes and differ from daily smokers with respect to key socioeconomic and psychological features[38]. Although we do not have nicotine dependency measures available for the current cohort, in our previous studies, we have shown that the Fagerstrom Test for Nicotine Dependence (FTND) is strongly correlated with the degree of cg05575921 demethylation while other have shown that FTND is associated with serum cotinine levels [21,39,40]. As a consequence, we believe that it may be possible to predict the intensity or subtype of addiction by determining the degree of demethylation. Conceivably, depending on how this smoking intensity maps on to the behavioural and physiological diatheses for smoking, this technique could have considerable influence on the cessation strategy used for a given patient.

Consistent with our prior findings, the simple linear relationship of cg05575921 methylation levels to serum cotinine levels was stronger than that of cg05575921 to self-report of cigarette consumption [21,24]. Part of this poorer fit of cg05575921 methylation to self-reported cigarette consumption could be due to our forcing subjects to choose one of seven consumption levels in this study. However, in our previous studies, we allowed subjects to exactly specify cigarette consumption and obtained similar results. From a strictly biological point of view, the stronger correlation of the biological measures with one another is to be expected. Both PAH and nicotine are absorbed into the bloodstream with

cg05575921 methylation and serum cotinine levels representing measures of PAH and nicotine catabolism, respectively. Furthermore, each of these two biomarker assays tap upon saturable biological pathways that have non-linear components [36,37]. Indeed, when a non-linear approach is used to model the relationship, the fit for each of these biomarkers to cigarette consumption markedly improves. Still, the reason for the less than perfect correspondence in the relationship between the amount of both cg05575921 and cotinine to the number of cigarettes consumed also may be secondary to the fact that the absorption PAH and nicotine is also dependent on several other variables including the number of puffs drawn from each cigarette, the puff volume of each cigarette, and the composition (filter, tobacco, etc.) of the cigarette itself. Because the hazards of smoking are directly related to the amount of toxins consumed, this may mean that serum cotinine or cg05575921 levels may be better predictors of adverse outcomes than self-reported measures. Indeed, in a number of analyses of the relationship of smoking to mortality, cg05575921 has been shown to be a substantially better predictor of death than self-reported intake [41,42].

These data replicate and extend prior findings that non-combustible consumption of tobacco, in the absence of other combusted substance use such as cannabis smoking, does not result in significant cg05575921 demethylation [20,21]. This is critically important for generating a better understanding of risks posed by vaping. Although vaping pens and e-cigarettes can generate PAHs, the amount of PAH generated from smoking an entire e-cigarette cartridge is 2 orders of magnitude lower than the amount of PAH generated from smoking a single cigarette (1 to 1.6 µg each) [43,44]. Next, second-hand exposure of PAHs from e-cigarettes results in even lower exposure to PAHs with a recent study showing that in a well-ventilated room, vaping only increased PAH levels only 20% above baseline. [45] This suggests that direct ingestion of limited amount of e-cigarette vapour or second-hand exposure to e-cigarette smoke will not noticeably affect DNA methylation at cg05575921 [45]. This statement should not be construed to be interpreted to mean that e-cigarette emissions are safe or do not affect DNA methylation elsewhere. Although comparatively low emitters of

PAH, e-cigarettes do emit significant levels of other environmental toxins [45]. Furthermore, one small study by Caliri and colleagues did show demethylation of LINE-1 elements in vapers [46]. But the study by Caliri and associates did not exclude those with histories of smoking more than 1 year prior to the study nor did it test for smoking in its vaping subjects. Given the 5–7% rate of unreliable self-report seen in this study as well as our prior studies and the failure of a well-powered genome-wide study by Besingi and Johansson to show any significant effects of snus use on DNA methylation, it is quite possible that the demethylation observed by Caliri may have been secondary to unobserved smoking behaviours. These discrepancies suggest that further examination of LINE-1 methylation in vaping subjects using biochemical validation of non-smoking status may be needed.

Although our findings that a carefully screened group of vapers do not show methylation patterns or mass spectroscopy metabolite patterns similar to smokers are novel, we were disappointed to find that propylene glycol (PG) levels did not distinguish those who vape from other subjects. Although PG is commonly used as a ‘humectant’ in all tobacco products, it is usually the major component of vaping fluid. Hence, we were hoping that much higher levels of PG would be found in the urine of our vapers as compared to the other use groups. Perhaps with greater numbers of subjects, a significant difference could have been found. Still, variability in the frequency of vaping combined with the short half-life of PG in the serum observed in this study may indicate that PG is a less than ideal biomarker for the objective confirmation of vaping status [47]. Overall, though, further work is needed to identify biomarkers that can reliably distinguish vaping from smokeless tobacco use.

In contrast, CEMA levels certainly seem capable of identifying subjects who use combustible forms of tobacco. Indeed, the fact that 3 of the 4 vapers with cg05575921 had CEMA levels above 3 ng/ml is strongly supportive of our suspicion that these subjects were surreptitiously smoking and demonstrates the value of these measurements for research purposes. Furthermore, our visual inspection of the anabasine and anatabine data along with the cg05575921 data from these same subjects have increased our suspicions that these self-reported

exclusive ENDS users may be surreptitiously smoking. Unfortunately, in our experience subjects do not like to provide urine samples and determining CEMA levels in urine using mass spectroscopy is very costly. Therefore, we regretfully conclude that it is unlikely that use of CEMA will become routine in studies of the relationship between ENDS use and smoking. But we firmly believe that using two good biomarkers of smoking is better than using just one.

Although the number of subjects is small, the finding that neither serum nor urinary cotinine levels were correlated with self-reports of vape pen usage highlights the fact that not all vaping solutions contain nicotine, and that among those that do, the concentration of nicotine found in them can vary significantly. [45,48] If the current findings are at least partially generalizable, they suggest that researchers seeking to understand the relationship of vaping to pulmonary related outcomes should not rely on cotinine levels to provide objective quantification of vaping solution consumption. In addition, these data may help explain the variability of the results of studies of the relationship of vaping to smoking and suggest a need for a more comprehensive understanding of the vaping fluid composition as a function of demographic and cultural variables [4,49–51].

We note as a limitation that the above subjects were carefully screened so that their specimens would reflect ‘pure’ patterns of nicotine use, i.e., smoking, vaping, and chewing. Because many individuals in the real world mix these different forms of use, and because co-use of combusted cannabis is common across each of these use groups, assessment of cg05575921 methylation may not always be an accurate indicator of whether an individual is or is not a cigarette smoker. For example, individuals who vape and regularly use combusted cannabis would be expected to demonstrate cotinine positivity as well as cg05575921 methylation below 80%. Second, we note that our examination of logistic regression models integrating different biomarkers as predictors of combusted versus non-combusted nicotine use is limited due to imbalanced sample sizes between use groups, which led us to combine the vaping and smokeless groups. This limitation stems from the termination of our study due to the COVID-19 pandemic, and suggests that future analyses with more balanced groups may refine

our estimates. Lastly, we note that, although cg05575921 performed well as a predictor of smoking status in combination with age, gender, and other smoking predictors, the inclusion of other CpGs associated with smoking may yield even stronger results in future studies.

As members of the scientific and medical communities, we hope that the approach delineated in this communication may help achieve a more comprehensive understanding of the effects of each type of nicotine product use. Specifically, we believe that the current findings which demonstrate that DNA methylation can reliably identify those subjects using combustible forms of tobacco will provide a method for others to better clarify the relationship of vaping to the onset of smoking and other forms of substance use. When used in combination with cotinine determinations, cg05575921 methylation should provide a powerful means of determining the nicotine or other related product use status. We are also hopeful that our other work exploring the capacity of DNA methylation to guide smoking cessation finds utility. Indeed, in the future, it very well may be that epigenetic approaches may find a prominent role in the assessment and management of tobacco use disorders.

Acknowledgments

We would like to thank Drs Max Scherer and Nikola Pluym of ABF Labs (Munich) for their kindness in discussing study design and their prompt conduct of our mass spectroscopy analyses.

Disclosure statement

Dr Philibert is the Chief Executive Officer of Behavioral Diagnostics. The use of cg05575921 to assess smoking status is covered by existing and pending patents including US Patents 8,637,652 and 9,273,358. The other authors have no applicable conflicts.

Funding

This work was supported by The Smoke Free World Foundation; Foundation for a Smoke Free World

Author contributions

AA, RR, and RP obtained the funding for the project. AA, RR, AB, NH, SM, and BH devised the clinical interview procedures, collected the clinical data, provided initial interpretation of clinical findings, and provided summary clinical variables. SM and KD performed the laboratory assessments and collaborated with AA, AB, NH, SM and BH in quality control measures. MD, JM, and JL analysed the data, composed all figures, and provided the bulk of the final draft of the results sections. RP wrote the initial manuscript draft with AA, MD, RR, JL, and JM providing additional material. KD provided proofreading services. All authors read and approved the final draft.

Consent

There are no images or details related to an individual person in this manuscript.

Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

Ethics

All protocols and procedures used in this study were approved by the University of Iowa Institutional Review Board (IRB 201,905,678). For this study, the University of Iowa has agreed to serve as the lead IRB for both the Iowa and the Des Moines University activities.

ORCID

Allan Andersen  <http://orcid.org/0000-0001-7006-7639>

References

- [1] Walley SC, Wilson KM, Winickoff JP, et al. A public health crisis: electronic cigarettes, vape, and JUUL. *Pediatrics*. 2019;143(6):e20182741.
- [2] Miech R, Johnston L, O'Malley PM, et al. Adolescent vaping and nicotine use in 2017–2018—US national estimates. *N Engl J Med*. 2019;380:192–193.
- [3] Dwyer JB, McQuown SC, Leslie FM. The dynamic effects of nicotine on the developing brain. *Pharmacol Ther*. 2009;122:125–139.
- [4] Chatterjee K, Alzghoul B, Innabi A, et al. Is vaping a gateway to smoking: a review of the longitudinal studies. *Int J Adolesc Med Health*. 2016;30(3):20160033.
- [5] Chaumont M, van de Borne, P., Bernard, A., et al. Fourth generation e-cigarette vaping induces transient lung inflammation and gas exchange disturbances: results from two randomized clinical trials. *Am J Physiol Lung Cell Mol Physiol*. 2019;316(5):L705–L719
- [6] Blount BC, Karwowski MP, Shields PG., et al. Vitamin E acetate in bronchoalveolar-lavage fluid associated with EVALI. *N Engl J Med*. 2019;382(8):697–705
- [7] Traboulsi H, Cherian M, Abou Rjeili M, et al. Inhalation toxicology of vaping products and implications for pulmonary health. *Int J Mol Sci*. 2020;21(10):3495
- [8] US. Department of health human services. The health consequences of smoking—50 years of progress: a report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease ..., 2014.
- [9] Creamer MR, Wang TW, Babb S, et al. Tobacco product use and cessation indicators among adults—United States, 2018. *Morbidity Mortality Weekly Rep*. 2019;68(45):1013
- [10] Florescu A, Ferrence R, Einarson T, et al. Methods for quantification of exposure to cigarette smoking and environmental tobacco smoke: focus on developmental toxicology. *Ther Drug Monit*. 2009;31(1):14–30
- [11] Jacob P, Hatsukami D, Severson H, et al. Anabasine and anatabine as biomarkers for tobacco use during nicotine replacement therapy. *Cancer Epidemiol Biomarkers Prev*. 2002;11:1668–1673
- [12] Schick SF, Blount BC, Jacob 3rd P, et al. Biomarkers of exposure to new and emerging tobacco delivery products. *Am J Physiol Lung Cell Mol Physiol*. 2017;313(3):L425–L452
- [13] Andersen AM, Philibert RA, Gibbons FX, et al. Accuracy and utility of an epigenetic biomarker for smoking in populations with varying rates of false self-report. *Am J Med Genet B Neuropsychiatr Genet*. 2017;174(6):641–650.
- [14] Gao X, Jia M, Zhang Y, et al. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin Epigenetics*. 2015;7(1):113.
- [15] Monick MM, Beach SR, Plume J, et al. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *Am J Med Genet B Neuropsychiatr Genet*. 2012;159B(2):141–151
- [16] Zeilinger S, Kühnel B, Klopp N, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS ONE*. 2013;8(5):e63812
- [17] Esser C. Biology and function of the aryl hydrocarbon receptor: report of an international and interdisciplinary conference. *Arch Toxicol*. 2012;86:1323–1329.
- [18] Benowitz NL, Swan GE, Jacob P III, et al. CYP2A6 genotype and the metabolism and disposition kinetics of nicotine. *Clin Pharmacol Ther*. 2006;80(5):457–467.
- [19] Wahlberg I, Ringberger T. Smokeless tobacco. In: D. Layten Davis, Mark T. Nielsen (eds.). *Tobacco production, chemistry and technology*. Blackwell Science; 1999.

- [20] Besingi W, Johansson Å. Smoke related DNA methylation changes in the etiology of human disease. *Hum Mol Genet.* **2013**;23(9):2290–2297.
- [21] Philibert R, Hollenbeck N, Andersen E, et al. A quantitative epigenetic approach for the assessment of cigarette consumption. *Front Psychol.* **2015**;6:656
- [22] Philibert R, Dogan M, Noel A, et al. Dose response and prediction characteristics of a methylation sensitive digital PCR assay for cigarette consumption in adults. *Front Genet.* **2018**;9. doi:10.3389/fgene.2018.00137
- [23] Dawes K, Andersen A, Vercande K, et al. Saliva DNA methylation detects nascent smoking in adolescents. *J Child Adol Psychopharm.* **2019**;29(7):535–544
- [24] Philibert R, Dogan M, Beach SRH, et al. AHRR methylation predicts smoking status and smoking intensity in both saliva and blood DNA. *Am J Genet.* **2019**;183(1):51–60.
- [25] Dawes K, Andersen A, Papworth E, et al. Refinement of cg05575921 demethylation response in nascent smoking. *Clin Epigenetics.* **2020**;12(1):1–11
- [26] Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform.* **2019**;95:103208
- [27] Hamilton CM, Strader LC, Pratt JG, et al. The phenx toolkit: get the most from your measures. *Am J Epidemiol.* **2011**;174(3):253–260
- [28] Philibert RA, Penaluna B, White T, et al. A pilot examination of the genome-wide DNA methylation signatures of subjects entering and exiting short-term alcohol dependence treatment programs. *Epigenetics.* **2014**;9(9):1212–1219
- [29] Luo X, Carmella SG, Chen M, et al. Urinary Cyanoethyl mercapturic acid, a biomarker of the smoke toxicant acrylonitrile, clearly distinguishes smokers from nonsmokers. *Nicotine Tob Res.* **2020**;22(10):1744–1747
- [30] Chang CM, Edwards SH, Arab A, et al. Biomarkers of tobacco exposure: summary of an FDA-sponsored public workshop. *Cancer Epidemiol Prev Biomarkers.* **2017**;26(3):291–302
- [31] Kramer CY. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics.* **1956**;12(3):307–310.
- [32] De Boor C, De Boor C, Mathématique E-U, et al. A practical guide to splines. Springer-Verlag New York; New York, **1978**
- [33] Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. Dordrecht, The Netherlands: D. Reidel. **1986**;81(10.5555)
- [34] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* **1983**;148(3):839–843.
- [35] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* **1988**;44(3):837–845.
- [36] Simon TW, Budinsky RA, Rowlands JC. A model for aryl hydrocarbon receptor-activated gene expression shows potency and efficacy changes and predicts squelching due to competition for transcription co-activators. *Plos One.* **2015**;10(6):e0127952.
- [37] Hukkanen J, Jacob P, Benowitz NL. Metabolism and disposition kinetics of nicotine. *Pharmacol Rev.* **2005**;57(1):79–115.
- [38] Schane RE, Glantz SA, Ling PM. Nondaily and social smoking: an increasingly prevalent pattern. *Arch Internal Med.* **2009**;169(19):1742–1744.
- [39] Pomerleau CS, Pomerleau OF, Majchrzak MJ, et al. Relationship between nicotine tolerance questionnaire scores and plasma cotinine. *Addict Behav.* **1990**;15(1):73–80.
- [40] Van Overmeire IP, De Smedt T, Dendale P, et al. Nicotine dependence and urinary nicotine, cotinine and hydroxycotinine levels in daily smokers. *Nicotine Tob Res.* **2016**;18(9):1813–1819
- [41] Zhang Y, Schöttker B, Florath I, et al. Smoking-associated DNA methylation biomarkers and their predictive value for all-cause and cardiovascular mortality. *Env Health Perspect.* **2015**;124(1):67–74
- [42] Bojesen SE, Timpson N, Relton C, et al. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax.* **2017**;72(7):646–653.
- [43] Ding YS, Trommel JS, Yan XJ, et al. Determination of 14 polycyclic aromatic hydrocarbons in mainstream smoke from domestic cigarettes. *Environ Sci Technol.* **2005**;39(2):471–478.
- [44] Cheng T. Chemical evaluation of electronic cigarettes. *Tob Control.* **2014**;23(suppl 2):ii11–ii17.
- [45] Schober W, Szendrei K, Matzen W, et al. Use of electronic cigarettes (e-cigarettes) impairs indoor air quality and increases FeNO levels of e-cigarette consumers. *Int J Hyg Environ Health.* **2014**;217(6):628–637
- [46] Caliri AW, Caceres A, Tommasi S, et al. Hypomethylation of LINE-1 repeat elements and global loss of DNA hydroxymethylation in vapers and smokers. *Epigenetics.* **2020**;15(8):816–829.
- [47] Landmesser A, Scherer M, Pluym N, et al. Biomarkers of exposure specific to E-vapor products based on stable-isotope labeled ingredients. *Nicotine Tob Res.* **2018**;21(3):314–322
- [48] Cameron JM, Howell DN, White JR, et al. Variable and potentially fatal amounts of nicotine in e-cigarette nicotine solutions. *Tob Control.* **2014**;23(1):77–78
- [49] Primack BA, Soneji S, Stoolmiller M, et al. Progression to traditional cigarette smoking after electronic cigarette use among US adolescents and young adults. *JAMA Pediatr.* **2015**;169(11):1018–1023.
- [50] Etter J-F. Gateway effects and electronic cigarettes. *Addiction.* **2018**;113(10):1776–1783.
- [51] Conner M, Grogan S, Simms-Ellis R, et al. Do electronic cigarettes increase cigarette smoking in UK adolescents? Evidence from a 12-month prospective study. *Tob Control.* **2018**;27(4):365–372